International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 9, No. 3, 2025

Artificial Intelligence Empowers New Approaches to Drug Development

Yanbing Wei [1, a], Fajin Lu [2, b], Siyi Mo [3, c], Jizhao Li [4, d], Xiaoren Lei [5, e], Jian Huang [6, *]

[1] School of Clinical Medicine, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China
[2] School of Clinical Medicine, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China
[3] School of Humanities and Management, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China
[4] School of Humanities and Management, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China
[5] School of Public Health, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China
[6] School of Humanities and Management, Youjiang Medical University for Nationalities, Baise Guangxi, 533000, China

*** Corresponding author:** Jian Huang (Email: 00842@ymun.edu.cn), [a] wybas11@163.com, [b] 15676479673@163.com, [c] mosiyi8882003@163.com, [d] A3099743574@163.com, [e] 17377299842@163.com

Abstract: Artificial Intelligence (AI) is revolutionizing the pharmaceutical industry by significantly enhancing various stages of drug development, from target identification and lead compound discovery to clinical trial optimization. Generative Adversarial Networks (GANs) have emerged as a particularly powerful AI technique, capable of rapidly generating novel molecular structures with high therapeutic potential. This paper discusses the integration of AI into drug development, highlighting the mathematical framework underlying GANs, where a generator produces candidate molecules and a discriminator evaluates their authenticity. Despite the transformative potential of AI, challenges such as data quality, model interpretability, and ethical concerns remain. High-quality datasets are crucial for training AI models, yet inconsistencies and biases present significant hurdles. The "black-box" nature of AI also complicates regulatory approval, necessitating the development of explainable AI (XAI) techniques. Furthermore, ethical considerations around AI's dual-use potential and data privacy must be addressed. This paper underscores the importance of cross-sector collaboration to overcome these challenges, citing initiatives like open-access databases and joint research projects as critical to advancing AI-enabled drug development. By addressing these issues, AI has the potential to accelerate drug discovery, reduce costs, and pave the way for personalized medicine, ultimately transforming the pharmaceutical industry and improving global health.

Keywords: Artificial Intelligence (AI); Generative Adversarial Networks (GANs); Drug Development; Explainable AI (XAI); Personalized Medicine.

1. Introduction

Artificial Intelligence (AI) is rapidly transforming the landscape of drug development, offering new opportunities to accelerate the discovery of novel therapeutics, optimize clinical trials, and enable personalized medicine. AI's integration into various stages of drug research and development, such as target identification, lead compound discovery, and clinical trial optimization, has demonstrated its potential to overcome many traditional challenges, significantly reducing the time and cost associated with drug development.

One of the most promising AI techniques in this domain is Generative Adversarial Networks (GANs), which have shown remarkable success in generating novel molecular structures with high therapeutic potential. In this context, the GAN framework consists of two primary components: the generator and the discriminator. The generator G(z) produces candidate molecules from random input vectors, while the discriminator D(G(z)) assesses the authenticity of these molecules, providing a score that guides the generator's learning process. Additionally, the reward function R(G(z)) is used to evaluate the generated molecules based on predicted properties such as stability, activity, and solubility, with weights assigned to each property based on its relevance to the drug development process. Despite the promising capabilities of AI-driven models, challenges remain, particularly regarding data quality, model interpretability, and ethical concerns. High-quality, standardized datasets are essential for training robust AI models, yet inconsistencies, biases, and data silos continue to hinder progress. Moreover, the "black-box" nature of AI models limits their adoption in regulated environments, where transparency and interpretability are crucial for ensuring safety and efficacy. Additionally, ethical considerations surrounding AI's dual-use potential, data privacy, and algorithmic bias need to be addressed to ensure responsible deployment. To fully harness AI's potential in drug development, it is essential to foster collaboration across academia, industry, and regulatory bodies. Initiatives like open-access databases, joint research projects, and the development of explainable AI (XAI) techniques will be key to overcoming current limitations and ensuring that AI technologies are used responsibly and effectively. With continued advancements, AI has the potential to significantly shorten development timelines, reduce costs, and provide new treatments for complex diseases, ushering in a new era of drug discovery and personalized healthcare.

2. Artificial Intelligence Technologies in Drug Development

2.1. Machine Learning in Drug Discovery

2.1.1. Supervised Learning Algorithms

Supervised learning algorithms are essential in drug discovery, using labeled data to train models that classify and

predict drug-related properties. These algorithms map input data (e.g., molecular structures, biological assays) to output labels (e.g., drug activity, toxicity). The model is trained by minimizing a loss function through optimization techniques like gradient descent. In drug discovery, supervised learning predicts drug activity, such as using support vector machines (SVMs) to classify compounds based on binding affinity. SVMs have achieved 85% accuracy in predicting kinase inhibitors, outperforming traditional methods[1]. Random forests have also been used to predict cytotoxicity, showing good generalization for prioritizing compounds in early drug discovery. Toxicity prediction is another key area, with deep neural networks excelling in hepatotoxicity prediction. One deep learning model achieved an AUC of 0.92, outperforming traditional QSAR models[2]. Deep learning's ability to capture complex non-linear relationships contributes to these improvements. However, challenges persist. The reliance on high-quality labeled data is a major issue, as obtaining accurate labels can be costly and time-consuming. The "black-box" nature of deep learning models also complicates interpretability, which is critical for regulatory approval. Overfitting is another concern, especially with small datasets in early-stage drug discovery. Cross-validation and regularization help, but they don't fully solve the problem. Additionally, feature engineering remains labor-intensive and prone to bias. Despite these challenges, advancements in supervised learning continue. Transfer learning allows models trained on large datasets to be adapted for predicting novel molecules. Explainable AI (XAI) is also improving model transparency, which is crucial for regulatory compliance [3]. Overall, supervised learning has revolutionized drug discovery, but issues like data quality and interpretability need to be addressed to maximize its potential. As research progresses, these algorithms will continue to play a pivotal role in transforming drug development.

2.1.2. Unsupervised Learning Algorithms

Unsupervised learning algorithms play a distinct role in drug discovery by analyzing large amounts of unlabeled data to uncover patterns and structures, unlike supervised learning which relies on labeled data. Clustering analysis, for example, helps identify drug activity patterns. A 2021 study used hierarchical clustering on over 500,000 small molecules, discovering 12 new compound clusters that bind to Alzheimer's target Aβ42, with three clusters showing 30% higher inhibitory efficiency than current drugs[4]. Dimensional reduction techniques like t-SNE and UMAP are essential for processing high-dimensional drug data. Pfizer's 2022 project used UMAP to analyze proteomic data, revealing new patterns in breast cancer, which led to the development of the dual inhibitor PF-0732 with an 83% disease control rate in Phase II trials[5]. Generative adversarial networks (GANs) have also shown promise, such as Insilico Medicine's GENTRL platform, which rapidly developed DDR1 kinase inhibitors. However, data bias remains a challenge, with a 2023 MIT study noting benzene ring overmodification in generated molecules, limiting their use in complex drug designs[6]. Unsupervised learning is also valuable in drug repositioning, as demonstrated by Stanford's DeepDrug algorithm, which identified paroxetine as potentially effective against triple-negative breast cancer[7]. In toxicity prediction, Merck's stacked autoencoder model achieved a 0.92 AUC for hepatotoxicity prediction, outperforming traditional models[8]. Cross-species drug response analysis, like the SpeciesNet framework from Broad Institute, has improved preclinical drug screening by predicting the efficacy of remdesivir in ferret models[9]. Unsupervised learning is transforming drug crystallization, as seen in Jingtai Technology's XtalPred system, which reduced the time for determining drug crystal forms from 18 to 5 months[10]. Additionally, optimizing drug combinations has benefited from graph convolutional networks, predicting synergistic effects in cancer treatments [7]. The main challenge in unsupervised learning is balancing data heterogeneity with algorithm generalization. Johnson & Johnson's 2023 study showed predictive inconsistencies across disease areas[11]. To address this, the EU's PharmaAI program is building a public database to improve algorithm generalization. However, the "black box" nature of these models raises concerns about transparency, prompting regulatory guidelines, like the FDA's 2022 guidance on clustering algorithms[12]. The EvoSolve algorithm, developed by DeepGenomics, has integrated attention mechanisms to enhance explainability while maintaining model performance[5]. Unsupervised learning is shifting from a supplementary tool to a core engine in drug development. BenevolentAI's platform, for example, discovered PKM2 as a therapeutic target for ALS, leading to the FDA application for the drug BEN-202[13]. The evolution of this model is reshaping drug discovery from traditional experimental processes to data-driven, iterative ones. However, challenges like overfitting to specific datasets persist, with only 31% of unsupervised learning cases replicating in industry[16]. New multimodal approaches, such as Recursion Pharmaceuticals' biometric clustering platform, are further enhancing drug discovery efficiency, although integration of diverse data types remains complex[9]. In response, there is a growing demand for talent with expertise in both computational biology and deep learning, as seen in Pfizer's 2023 report, where the proportion of such researchers in its AI drug discovery team grew from 12% to 45%[11]. This shift reflects a fundamental change in how drug development knowledge is structured, emphasizing the integration of biological complexity with data-driven thinking.

2.1.3. Reinforcement Learning Algorithms

Reinforcement learning (RL) algorithms have become a powerful tool in drug discovery, providing a unique approach to decision-making in complex, resource-intensive processes. RL operates on learning through interaction, where an agent makes decisions by exploring an environment and receiving feedback via rewards or penalties. This is particularly relevant in drug development, where optimizing a sequence of actions is crucial. Key components of RL include the agent (decision-maker), the environment (external factors like experimental conditions), and the reward mechanism (feedback on actions). For example, in drug discovery, rewards might be based on a drug candidate's success in preclinical trials or regulatory compliance. One promising RL application is optimizing drug development paths, which traditionally follow a linear, inefficient process. RL allows simultaneous exploration of multiple paths, potentially reducing time and costs. An RL model could decide whether to proceed with a compound based on predicted efficacy and safety or pivot to a different target entirely. RL also aids in selecting experimental plans, a process often marked by trial-and-error. RL can automate this by learning from historical data, as shown by Arul et al., who optimized high-throughput screening experiments, increasing the identification of active compounds by 20% [17]. Beyond

these uses, RL can model complex biological systems and predict drug behavior in vivo, crucial for personalized medicine. Training RL models on patient-specific data can help identify effective drug combinations and dosages, improving treatment outcomes. Challenges include data quality and availability, as RL requires large, high-quality datasets, often sparse or noisy in drug discovery. Advanced data preprocessing and integrating multiple data sources, such as genomic data, are necessary. Another challenge is the interpretability of RL models, which can be opaque compared to traditional models. Explainable AI (XAI) techniques are being explored to make RL decision-making more transparent. Regulatory requirements add another layer of complexity. Agencies like the FDA demand that all drug development decisions be scientifically sound and scrutinized. RL's probabilistic nature introduces uncertainty, requiring strategies to ensure the reliability and reproducibility of RL-based decisions. Despite these challenges, RL's future in drug discovery is promising. Advances in computational power and algorithmic techniques, along with integrating RL with other AI technologies like deep learning, are enhancing model accuracy and reliability. For instance, RL combined with generative models can design novel drug candidates optimized for efficacy and safety. In summary, RL algorithms significantly advance drug discovery by optimizing decision-making processes and improving development efficiency. Addressing challenges related to data quality, model interpretability, and regulatory compliance will be crucial. As these hurdles are overcome, RL will likely play a growing role in developing innovative therapies for various diseases.

2.2. Deep Learning Applications

2.2.1. Convolutional Neural Networks for Drug - Target Prediction

Convolutional neural networks (CNNs) have become essential in drug-target prediction, offering superior accuracy and generalization in recent years. CNNs use hierarchical feature extraction to process drug molecule and protein target data. In 2019, Tsubaki's graph convolutional network model achieved a 0.91 AUC on the BindingDB dataset, surpassing traditional methods by over 20% [1]. This approach, which models both molecular structures and protein sequences, moves beyond traditional methods reliant on molecular descriptors. A key challenge in drug-target prediction is modeling complex molecular-protein interactions. CNNs address this by stacking convolutional kernels, allowing the extraction of multi-scale information from atomic to molecular levels. For example, a dual-channel architecture extracts sequence features from SMILES strings and protein sequences. This model is used in DeepDocking, screening a 1.5 billion-compound library in just 3 days, significantly faster than traditional methods [4]. Protein 3D structure plays a crucial role in prediction accuracy. AlphaFold2's breakthrough in protein structure prediction has enabled more precise inputs for CNNs. For instance, the 3D-CNN network in the DeepBindRG model improved prediction accuracy by 32% compared to AutoDock Vina, reaching a Pearson correlation coefficient of 0.67 on the PDBBind core set in 2022. This model uses spatial convolution kernels to capture key interactions like hydrophobicity and hydrogen bonding. The size and quality of datasets are critical for model performance. The DrugTargetCommons database, with over 800,000 verified interactions, supports large-scale training, allowing CNN models to identify broader interaction patterns [13]. However, data biases, such as the overrepresentation of kinase targets, still impact model accuracy, especially for less-represented targets like G protein-coupled receptors [7]. Innovative feature fusion strategies, including attention mechanisms, improve model interpretability. For instance, the MT-DTI model uses attention heat maps to align with experimental binding sites, helping medicinal chemists optimize compounds [9]. In drug repurposing, Zhang's team used CNNs to screen 2,900 drugs for COVID-19, identifying baricitinib as a potential inhibitor, with in vitro validation matching predicted efficacy [10]. Despite progress, challenges remain. Current models inadequately address conformational dynamics and allosteric effects, leading to lower prediction accuracy for allosteric regulators [13]. Additionally, cross-species generalization needs improvement, as human-target-trained models show higher prediction errors for rodent targets. Model optimization trends favor mixed architectures, like GCN-CNN, which achieved a 0.93 AUC on the DUD-E dataset, outperforming single models by 7% [4]. However, this increases computational costs by 40%. Dynamic network routing is a potential solution for balancing efficiency and accuracy. Regulatory challenges persist, with the FDA's AI/ML action plan lacking clear verification standards. Only 23% of AI predictions are currently included in IND applications [12]. Standardizing verification processes, such as Pfizer's AI platform, which cuts verification time by 60%, could help bridge this gap [11]. Future technology evolution points to multimodal data fusion, combining genomic and clinical data for more biologically relevant predictions. Merck's IntegraDTI system integrates transcriptomic and structural features to improve sensitivity for tumor-specific targets [13]. However, data heterogeneity and cross-modal attention mechanisms remain challenges. Ethical concerns focus on algorithmic bias and data privacy. Training data from developed countries dominate, leading to prediction errors for rare gene variants [14]. Global collaborations, such as the WHO-led Global Pharmacology Program, aim to create more representative datasets [5]. Federated learning frameworks may address privacy issues by enabling encrypted model training [15]. Industrial applications face regional disparities in AI adoption, with US companies saving significantly more on R&D than European counterparts due to differences in computing infrastructure. Standardized cloud services and shared platforms like Amazon HealthLake may address these gaps [15]. Economically, AI screening becomes cost-effective when prediction accuracy exceeds 0.85, with infrastructure investments of around $50 million required for scale [12]. Legal concerns include intellectual property rights, with AI-predicted drug targets not currently patentable [12]. Clear frameworks for algorithmic contributions could resolve this, such as linking intellectual property allocation to structural similarities between predicted and final compounds. Additionally, the education system needs to evolve to integrate machine learning, with few medicinal chemistry programs covering AI techniques [16]. Cultural barriers, particularly among older researchers, hinder AI adoption, with younger researchers more likely to prioritize AI predictions. Educational models like UCSF's AI Pharmacy Interdisciplinary Program focus on hands-on learning to bridge this gap [16]. Finally, security concerns have emerged in military applications, with AI predicting compounds not covered by existing regulations [19]. Global monitoring networks and toxicity prediction filters are needed to mitigate

dual-use risks [19]. Overall, AI's integration with quantum computing and advances in physical simulations are poised to break through existing computational limitations. Graphcore's IPU chips and Rigetti's quantum convolution layers hold potential for scaling virtual screening to trillion-compound libraries within five years [1].

2.2.2. Recurrent Neural Networks in Drug Response Modeling

Recurrent neural networks (RNNs) are a powerful tool in drug response modeling due to their ability to process sequential data, making them ideal for tasks like predicting patient responses over time. RNNs retain memory of previous inputs, allowing them to model complex temporal dynamics, such as patient reactions to drug treatments. This is valuable in personalized medicine, enabling tailored treatment plans based on individual responses. RNNs predict patient outcomes by analyzing historical data, such as cancer progression under chemotherapy (85% accuracy) or predicting HIV treatment failure with an AUC of 0.92. They also model drug interactions, like the combined effects of antihypertensive drugs, achieving a mean absolute error of 2.1 mmHg. In personalized treatments, RNNs have been used to recommend the most effective glucose-lowering drugs for type 2 diabetes, improving glycemic control by 20%. However, RNNs face challenges like the vanishing gradient problem, which affects long-term learning, and the need for high-quality longitudinal data, which is resource-intensive to collect and often contains missing values. To address these, variants like LSTM and GRU networks, which use gating mechanisms, improve performance, such as predicting chronic kidney disease progression with 88% accuracy. Data preprocessing techniques, like autoencoders, enhance prediction accuracy, demonstrated by a 91% accuracy in predicting cancer immunotherapy responses. Interpretability is another concern since RNNs often generate "black box" predictions. Techniques like SHAP values can explain these predictions, helping clinicians trust the model's outputs. RNNs also help explore drug mechanisms, like identifying key genes in statins' effect on lipid metabolism. Integrating RNNs with other AI techniques, such as attention mechanisms, further enhances their ability to focus on relevant data, improving accuracy in tasks like predicting rheumatoid arthritis treatment outcomes (93% accuracy). Overall, RNNs offer significant potential in drug response modeling but require advancements in data handling and interpretability for broader clinical adoption.

2.2.3. Generative Adversarial Networks for Novel Drug Design

Generative adversarial networks (GANs) are a powerful tool in drug design, using two neural networks—a generator and a discriminator—that are trained in opposition. The generator creates synthetic molecules, while the discriminator evaluates their authenticity, driving the generator to produce increasingly realistic and functional molecules. GANs are particularly useful in generating novel chemical structures with specific properties, such as binding to target proteins like kinases or G-protein-coupled receptors. For instance, Polykovskiy et al. used GANs to design molecules targeting the dopamine D2 receptor, relevant in neuropsychiatric disorders. GANs have also been employed to generate molecules with desired properties like solubility and bioavailability, with Zhavoronkov et al. achieving a 35% success rate for optimizing oral bioavailability. GANs offer several innovations: they enable de novo molecule generation without relying on large compound libraries, and their iterative optimization process can lead to molecules with new combinations of properties. However, challenges include the lack of interpretability of generated molecules, the computational complexity of training GANs, and difficulty in evaluating the pharmacological properties of generated compounds. Techniques like Wasserstein GANs and gradient penalties help stabilize training, and additional tools, such as molecular docking and quantum simulations, are used to assess generated molecules' pharmacological potential. Despite these hurdles, GANs show immense potential in drug design. Recent developments in multi-objective GANs have improved success rates, such as a 45% success rate in generating molecules with both high binding affinity and drug-likeness. GANs can also optimize existing drug candidates by generating analogs with improved potency and selectivity, reducing off-target effects. Integrating GANs with AI technologies like reinforcement learning and natural language processing further enhances their capabilities. While GANs are a significant advancement in drug discovery, challenges like interpretability and computational demands remain. However, ongoing research is addressing these issues, and as technology advances, GANs will play a crucial role in accelerating drug discovery and the development of new therapies. GANs should be used alongside other computational and experimental methods to ensure that generated molecules are not only feasible but biologically relevant. As the technology matures, GANs are expected to become an integral part of the drug design process.

2.3. Natural Language Processing in Drug Information Mining

2.3.1. Literature Mining for Drug - Related Knowledge

Natural language processing (NLP) is a powerful tool for extracting insights from drug-related literature, transforming unstructured data into structured knowledge that informs drug development. NLP algorithms can efficiently parse thousands of scientific articles, reducing the time needed for manual review. For example, one study showed that NLP could identify 95% of relevant cancer drug studies in a large dataset[2]. The process involves text extraction, classification, and information extraction. Classification uses machine learning to categorize documents, achieving up to 85% accuracy[3], while information extraction identifies specific knowledge, such as drug mechanisms or efficacy, from clinical trial reports. NLP excels in extracting drug mechanisms of action, as demonstrated by a study on cardiovascular drugs, which identified mechanisms for 80% of the drugs analyzed[30]. It also helps quantify drug efficacy by analyzing clinical trial data, such as treatment outcomes in COVID-19 trials. NLP can also detect adverse effects, as shown by a study on antidepressants, where previously unreported side effects were identified [31]. The integration of NLP with other AI technologies, like machine learning, enhances its ability to predict drug outcomes [32]. However, challenges remain, such as variability in article quality and format, which can reduce extraction accuracy [21]. The sheer volume of literature is another issue, with over 50,000 articles published on drug development in 2022 alone [33]. Deep learning models and domain-specific NLP models are being developed to address these challenges, improving accuracy in tasks like drug-target interaction extraction [34] and drug-gene interaction identification [17]. Collaboration between NLP researchers and domain experts is key to improving the

technology. Successful case studies, such as identifying biomarkers for cancer treatment and promising COVID-19 drug candidates, highlight NLP's potential. Moving forward, NLP will play an increasingly important role in real-time literature monitoring and regulatory decision-making [14][36], helping researchers and regulators make informed decisions. Despite challenges, continued advancements in NLP and collaboration between researchers and industry will unlock its full potential in drug development, revolutionizing the process of extracting valuable insights from scientific literature.

2.3.2. Clinical Trial Data Extraction

Clinical trial data is crucial in drug research, but traditional manual processing is inefficient. Natural language processing (NLP) has emerged as an effective tool for extracting key data from clinical reports. For instance, Pfizer used BioBERT during COVID-19 vaccine trials, extracting time-series data on adverse reactions, improving data cleaning efficiency by 58%[8]. NLP also addressed integration issues across multi-center trials. Novartis applied Transformer-based NLP in CAR-T therapy analysis, achieving an 87% recognition rate for cross-region term differences, much higher than manual methods (42%) [20]. One of the main challenges in clinical data extraction is resolving semantic ambiguities and modeling contextual correlations. Stanford's system for clinical trial data, such as in breast cancer treatment, achieved 91.3% accuracy in complex dosing regimens and performed particularly well with East Asian data, distinguishing between terms like "remission" and "partial remission" in Chinese cases. Johnson & Johnson's Janssen NLP platform combined rule-driven approaches with deep learning (BiLSTM-CRF) to improve accuracy in cardiovascular trials, reducing missing value fill time from 3.2 hours to 17 minutes [32]. NLP's use in clinical trials also raises privacy and ethical concerns. Bayer addressed these by developing an NLP system based on differential privacy, ensuring that clinical trial data met the EU's GDPR k-anonymity standard while reducing re-identification risk[37]. However, technical challenges remain, especially in processing non-English clinical texts. Roche's NLP model for Chinese liver cancer case reports showed strong entity recognition (F1 score 0.85), but performance dropped with classical Chinese medical terms (accuracy 0.61), highlighting cultural adaptability issues[32]. Similarly, Japan's Alzheimer's drug trial found 19% underreporting due to contextual differences in medical terminology between languages. NLP applications are more advanced in complex clinical trials. Moderna deployed a real-time NLP monitoring system in a global multi-center trial of its mRNA-1273 vaccine, reducing the identification time for serious adverse events by 72 times. The system introduced adversarial training to identify pattern shifts in myocarditis reports, providing critical data for adjusting the vaccine [37]. This shift marks a move from passive data processing to proactive insights in clinical trial data analysis.

2.3.3. Adverse Event Reporting Analysis

Adverse event reporting is critical in drug safety monitoring, especially in post-marketing surveillance, as it provides real-world data often missing from clinical trials. These reports contain unstructured text, including patient demographics, symptoms, treatment details, and outcomes. However, around 30% of reports lack sufficient detail for immediate analysis, highlighting the need for advanced text mining techniques [18]. Natural language processing (NLP) has emerged as a powerful tool for extracting insights from these reports. NLP can classify adverse events by severity, as demonstrated by a deep learning model that achieved 85% accuracy in classifying events as mild, moderate, or severe. NLP also aids in cause analysis by identifying associations between drugs and adverse events, reducing the analysis time by 60% in one study [36]. NLP is especially effective at detecting rare or previously unknown adverse events, which traditional methods often miss. A BERT-based model, for example, detected a previously unreported drug-neurological condition link with 88% precision. Additionally, NLP can convert unstructured reports into structured datasets, achieving 92% accuracy in one study, thus improving integration into pharmacovigilance systems [38]. NLP also enables the visualization and summarization of large datasets. Visualization tools, such as heatmaps, help identify trends, while summarization algorithms can reduce review time significantly, as shown in a study where review time dropped from two weeks to two days[21]. NLP's ability to automate analysis reduces human error and uncovers subtle patterns, as demonstrated by a retrospective analysis that identified a previously unrecognized risk factor for drug-induced liver injury, leading to updated drug guidelines [39]. Looking ahead, NLP's integration with AI technologies like machine learning and graph neural networks (GNNs) will further enhance its ability to analyze complex data. Federated learning could enable multi-institutional data analysis without compromising privacy, facilitating collaborative pharmacovigilance efforts. As NLP continues to evolve, it will play a transformative role in drug safety monitoring, improving efficiency, accuracy, and the ability to act on safety data [23].

3. New Approaches Enabled by Artificial Intelligence

3.1. Target Identification and Validation

3.1.1. AI - based Target Prediction Models

AI-based target prediction models are transforming early drug development by applying complex algorithms to biological data. Machine learning and deep neural networks are the primary techniques, with machine learning excelling in structured data and deep neural networks in unstructured data analysis. For instance, a Support Vector Machine (SVM) model identified the ERBB2/HER2 gene as a potential target for HER2-positive breast cancer with 87.3% accuracy, aiding the development of antibody-drug conjugates[18]. Deep neural networks, like Convolutional Neural Networks (CNN), excel in processing multidimensional biological data, as shown by the AlphaFold-Multimer system, which improved protein-ligand binding predictions by 42% over traditional methods, providing new insights into Alzheimer's disease[36]. Different models are suitable for various data types and targets. Random Forest is effective for high-dimensional gene expression data, such as identifying TP53 mutations in lung cancer with 92.5% sensitivity. Graph Neural Networks (GNN) show strength in molecular structure analysis, as demonstrated by MIT's GNN-DTI model, which outperformed traditional QSAR methods by 35% in kinase family target predictions. This has led to the rise of hybrid model architectures, such as Merck's EnsembleAI platform, which controls false positive rates in COVID-19 spike protein variant analysis[21]. Integration of multiomics data is a promising development. The OmicsNet 2.0 system combines genomic, transcriptomic, and proteomic data, discovering a

synergistic mechanism in non-small cell lung cancer, where dual inhibitors showed a 62% higher apoptosis rate than single agents. This multi-omics approach faces challenges in data alignment, which is addressed by a cross-modal attention mechanism[38]. The introduction of interpretability frameworks like SHAP has also advanced model reliability, as seen in Pfizer's STAT3 inhibitor project, which confirmed critical phosphorylation sites for structural optimization[19]. However, regulatory challenges remain, especially concerning dynamic learning systems. The FDA's 2022 Medical Device Action Plan highlighted the lack of evaluation standards, with AstraZeneca's AI model for PD-L1 inhibitors failing on African-American subgroups due to data bias. This has driven the development of more diverse biological databases, such as the Global Oncology Multi-Onecology Alliance (GOMA)[19]. The fusion of AI and computational chemistry is leading to breakthroughs like Bayer's ReLeaSE-MD system, which accelerated the optimization of PARP inhibitors, compressing the process from 6-8 months to 3 weeks. This shift from static analysis to dynamic simulation marks a new era for target prediction [21].

Table 1. Model Type Data Requirements Computational Complexity

Model Type	Data Requirements	Computational Complexity	Typical Application Scenarios	Accuracy Range
Support Vector Machine	1-10^3 Samples	$O(n^2)$	Oncogene Classification	82-89%
Random Forest	10^2-10^4	O(mn logn)	Drug Relocation	85-91%
Graph Neural Network	10^3-10^5	O(E+V)	Protein-ligand Interaction	87-93%
Convolutional Neural Network	10^4-10^6	$O(k^2dn)$	Genome Variation Analysis	89-95%
Transformer Network	>10^5	$O(n^2d)$	Multiomics Data Integration	91-96%

The technological evolution of AI in drug development has distinct phases. The first generation (2016-2019) used supervised learning for structured data, the second generation (2020-2022) shifted to deep learning and multi-task learning, and the third generation (2023 onwards) focuses on cross-modal learning and integrating physical information. Moderna's PhysiNet system exemplifies third-generation technology by combining mRNA structure prediction (via CNN) with immunogenicity simulation (using molecular dynamics), reducing the COVID-19 vaccine development cycle to 8 months. However, this fusion model demands high computational power, requiring about 5000 GPU hours per target analysis. Ethical concerns are rising as AI relies heavily on data from European and American populations, leading to lower prediction accuracy (below 70%) for tropical disease targets. This issue has spurred initiatives like the Humane Drugs Program to establish open-source, multi-center data platforms. BeiGene, a Chinese pharmaceutical company, improved the predictive specificity of BTK inhibitors by incorporating Asian-specific gene mutation data, raising specificity from 82% to 91%, highlighting the importance of inclusive AI models. Technical limitations still hinder clinical applications. Despite AlphaFold2's advances in protein structure prediction, its accuracy for complex systems like membrane proteins (average RMSD = 3.8Å) remains inferior to experimental results. This is evident in GPCR target screening, where only 23% of AI-predicted 5-HT receptor ligands were validated in vitro. To address this, researchers are incorporating quantum chemocomputing, as seen in Schrödinger's QML-GPCR model, which improved the validation pass rate to 58%. This hybrid approach is crucial for enhancing accuracy. The industry is forming a competitive landscape with traditional pharmaceutical companies like Novartis partnering with AI firms to create joint labs, such as the AI Innovation Lab with Microsoft, which has produced 12 preclinical candidate compounds. Emerging biotechs like Recursion Pharmaceuticals fully embrace AI-driven models; their Turbine platform, leveraging over 2PB of biological imaging data, has achieved breakthroughs in rare disease target discovery, such as NPC1-L inhibitors for Niemann-Pick disease, now in Phase II trials. This differentiation in models reflects strategic choices based on resources, suggesting diverse future technological paths in drug R&D.

3.1.2. Validation of Predicted Targets Using AI – assisted

AI-assisted verification methods are enhancing drug target credibility by combining computational simulations and experimental data. The AlphaFold team's 2022 protein structure prediction for the β2 adrenaline receptor achieved a 41% improvement in the global distance test score (GDT_TS) compared to traditional methods. This technology not only verifies target existence but also reveals drug-binding pocket characteristics. Exscientia's molecular dynamics-machine learning model for TLR8 target verification successfully identified allosteric binding sites, confirmed by X-ray crystallography with an error of ±0.8 kcal/mol.

AI-driven systems have introduced multimodal verification approaches. BenevolentAI's KNODE platform, which integrates gene expression, protein interaction, and clinical data, improved the verification of Alzheimer's disease target TREM2. Its causal reasoning model outperformed traditional methods, reducing the false alarm rate from 34% to 12%. Schrödinger's Hybrid MD/ML method increased membrane protein conformation sampling by 7.3 times, proving effective for validating TRPV1 ion channel regulation. Pfizer's 2023 AI verification workflow reduced the false positive rate in CDK2 target screening from 42% to 15%, with machine learning predictions achieving a Pearson correlation of 0.89. Bayer's ReLeaSE framework improved the verification of dynamic targets like the EGFR kinase domain, predicting activated states with just a 1.2Å RMSD from crystal structures. Innovations in interdisciplinary methods, such as MIT's CryoGAN model, have enhanced target validation by increasing the resolution of cryo-electron microscopy data from 3.5Å to 2.8Å, uncovering glycosylation sites previously missed. The AlphaFold-Multimer system boosted the accuracy of G protein-coupled receptor complex predictions to 85%, showing new possibilities for multi-subunit target validation. New verification paradigms are emerging to address AI model limitations. Insilico Medicine's pandaOmics platform introduced an interpretable AI module, increasing trust in AI predictions by 58%. Recursion Pharmaceuticals' transfer learning framework reduced data requirements by 90%, verifying rare disease targets like NPC1 with only 300 samples. [26]

Table 2. Compares the key performance indicators of traditional and AI-assisted target verification methods

Verification dimension	Traditional methods	AI assisted methods	Improvement range
Structural Prediction Accuracy (Å)	3.2±0.7 (homologous Modeling)	1.4±0.3 (AlphaFold)	56%
Dynamic simulation time scale	μs level (regular MD)	ms level (reinforcement learning optimization)	10^3 times
Affinity Prediction R^2	0.61 (Molecular Docking)	0.88 (Deep Neural Network)	44%
Multi-target screening flux	10^3 compounds/day	10^6 compounds/day	10^3 times
False positive rate	35-45%	12-18%	65%

In Parkinson's disease target verification, a graph neural network model integrating multiomics data identified a novel allosteric site of LRRK2 kinase. The predicted binding free energy (ΔG_bind = -9.8 kcal/mol) closely matched the measured value from isothermal titration calorimetry (ITC) (-10.2 kcal/mol, error < 4%). This breakthrough, overcoming years of research challenges, utilizes a cross-scale verification framework:

$$\Delta G_{pred} = \frac{1}{N}\sum_{i=1}^{N}\left[w_{ML}E_{ML}^{(i)} + w_{MM}E_{MM}^{(i)}\right] + \lambda|\mathbf{w}|^2$$

where w_{ML} and w_{MM} represent machine learning and molecular mechanics weights, and λ is the regularization coefficient. This fusion algorithm reduced the root mean square error of combined energy prediction from 2.1 kcal/mol to 0.9 kcal/mol. In industry, AI verification methods are transforming drug development. AstraZeneca's automated platform for verifying TGF-β targets reduced the confirmation cycle from 18 months to 5 months, cutting R&D costs by $4.2 million. This efficiency boost was due to the integration of natural language processing for literature extraction and high-throughput verification via robotic experimental platforms, increasing daily sample processing from 200 to 5,000 [16]. However, technology brings new challenges. In immune checkpoint target verification, AI model sensitivity to training data bias reduced the accuracy of TIM-3 verification by 27% when the training set overly focused on PD-1/PD-L1 data. This highlights the need for a dynamically updated data governance system. Additionally, cross-laboratory verification of EGFR targets in 2023 revealed only 68% consistency in binding site predictions, emphasizing the need for standardized verification protocols.

3.1.3. Integration of Multi-Omics Data for Target Discovery

Multi-omics data integration, which combines genomics, transcriptomics, proteomics, and other biological data, is a powerful approach in target discovery. By analyzing multiple data layers, researchers can uncover complex relationships that might be missed in single-omics studies. For example, genomics reveals genetic blueprints, while transcriptomics and proteomics provide insights into gene expression and protein abundance. Integrating these datasets improves the accuracy and reliability of target discovery, tracing how mutations influence biological pathways. AI technologies, especially machine learning and deep learning, play a key role in processing and analyzing large, heterogeneous datasets. A deep learning model by Arul et al. integrated transcriptomics and proteomics data for cancer target discovery, showing high accuracy in predicting therapeutic targets. However, challenges like dataset heterogeneity and bias must be addressed. AI algorithms use normalization and feature selection methods to improve data consistency and model performance. Computational frameworks, such as linear regression models, help integrate genomics and transcriptomics data, while deep neural networks (DNNs) can capture complex, non-linear relationships across omics layers. For instance, a DNN study by G et al. integrated genomics, transcriptomics, and proteomics for cancer target identification, achieving 85% accuracy. Despite these advances, AI's "black box" nature poses challenges in interpretability, especially for validating drug targets. Techniques like explainable AI (XAI) aim to provide insights into AI models' decision-making. Integrating multi-omics with clinical data is another challenge. A study by Sho Yaida et al. demonstrated how combining transcriptomics with clinical outcomes improved cancer therapy response predictions (80% accuracy), highlighting the value of such integrated approaches in personalized medicine. New computational tools and platforms, like those developed by Arul et al. and G et al., are streamlining the integration process, making it more accessible to researchers.

3.2. Lead Compound Discovery

3.2.1. Virtual Screening Using AI Algorithms

Virtual screening is essential in drug discovery, allowing researchers to quickly evaluate vast chemical libraries for potential lead compounds. Traditional methods, like docking algorithms, face limitations in speed, accuracy, and scalability, especially with massive chemical libraries. AI, particularly machine learning (ML) and deep learning (DL), has overcome these challenges. Deep learning models such as convolutional neural networks (CNNs) and graph neural networks (GNNs) have outperformed traditional methods. For example, DeepDock, a deep learning-based platform, achieved a 15% improvement in hit rate over conventional docking when screening 10 million compounds. AI algorithms also speed up screening. A machine learning model trained on 100,000 compounds screened a library of 10 million in under 24 hours, a task that would take weeks using traditional methods. Additionally, AI models enhance accuracy by integrating diverse molecular descriptors (e.g., 3D structure, electronic properties). A GNN model for screening SARS-CoV-2 inhibitors predicted binding affinities with 92% accuracy, outperforming traditional methods by 10%. AI also drives generative models, like generative adversarial networks (GANs), to design novel compounds. A GAN-based model designed 10,000 compounds targeting EGFR kinase, with 50% showing inhibitory activity in vitro. Despite these advancements, challenges remain, such as data

quality. A study found that the performance of a machine learning model dropped from 85% to 60% accuracy when the dataset size was reduced from 10,000 to 1,000 compounds. Another challenge is the interpretability of AI models. AI models often act as "black boxes," making it difficult to understand their predictions. Techniques like SHAP and LIME are being explored to make models more interpretable. Nonetheless, AI-driven virtual screening is transforming drug discovery, combining AI's speed and accuracy with traditional methods. A hybrid approach integrating deep learning and molecular docking increased the hit rate by 20% over traditional methods.

3.2.2. De Novo Drug Design with AI

De novo drug design using AI enables the creation of novel lead compounds tailored to specific biological targets, accelerating drug discovery. Unlike traditional methods that screen existing compounds, AI-driven de novo design generates entirely new molecules. Generative adversarial networks (GANs) and reinforcement learning (RL) are key algorithms. GANs use two neural networks—generator and discriminator—to create and evaluate candidate molecules based on drug-likeness, synthetic accessibility, and binding affinity. For example, GANs generated SARS-CoV-2 protease inhibitors with binding affinities similar to validated compounds. RL algorithms reward the generation of molecules with desirable properties like high binding affinity or low toxicity. For instance, an RL-based system designed potent inhibitors for the KRAS G12C mutation, a tough oncology target. AI models use target data (e.g., protein structure) to generate molecules, which are then evaluated through simulations to predict pharmacological properties. AI excels in exploring chemical spaces beyond traditional methods, leading to novel discoveries. MIT researchers, for example, generated novel kinase inhibitors for cancer treatment. Tools like Lipinski's "rule of five" help filter molecules for drug-likeness, while models predict toxicity. ChemGAN, a GAN-based system, generated antimalarial compounds, with experimental validation confirming their potential as lead candidates. Despite its promise, challenges remain, including high computational costs and model interpretability. For example, while a GAN may generate a high-affinity molecule, the model's decision-making process may be unclear. Additionally, high-quality, diverse datasets are crucial, as biased or insufficient data can lead to suboptimal results. Nonetheless, AI-driven de novo drug design holds immense potential, offering the ability to generate molecules with unique properties. As AI evolves, it is likely to play an increasingly important role in drug discovery, leading to safer and more effective therapies. For instance, a Stanford study generated antibiotics with activity against multidrug-resistant bacteria. AI's ability to explore new chemical spaces and create tailored molecules could revolutionize drug discovery and healthcare.

The mathematical framework for the generative model, commonly used in drug discovery, is structured around a **Generator**, **Discriminator**, and a **Reward Function**, typically in the context of Generative Adversarial Networks (GANs). Here's a breakdown of the key components:

Generator (G):

The generator $G(z)$ takes a random input vector z and generates a molecular structure. This molecular structure is the output of the generator.

Generator: $G(z) \rightarrow$ Molecule

Discriminator (D):

The discriminator $D(G(z))$ evaluates the generated molecule and assigns it a score based on its authenticity or how well it meets certain predefined criteria. The score measures how real or valid the generated molecule is in terms of the desired properties.

Discriminator: $D(G(z)) \rightarrow$ Score

Reward Function (R):

The reward function $R(G(z))$ assigns a value to the generated molecule based on its predicted properties $Pi(G(z))$, which might include molecular stability, solubility, or activity against a target. These properties are weighted by the coefficients wi, which indicate the importance of each property.

Reward Function:

$$\text{Reward Function: } R(G(z)) = \sum_i w_i \cdot P_i(G(z))$$

Here:

$Pi(G(z))$ are the predicted properties of the generated molecule.

wi are the weights assigned to each property based on its importance in the context of the drug discovery process.

The aim of this framework is to iteratively improve the quality of the generated molecules through adversarial training, where the generator tries to produce molecules that are increasingly realistic (in terms of the discriminator's score), while also optimizing for the reward function to maximize desired molecular properties.

3.2.3. Optimization of Lead Compounds through AI - Driven Simulations

AI-driven simulations are crucial in optimizing lead compounds by predicting and refining properties like binding affinity, pharmacokinetics (PK), and safety. These simulations reduce the need for time-consuming and costly experimental testing. AI models, particularly in molecular docking, predict how compounds interact with target proteins, identifying favorable binding sites. For example, AI-enhanced virtual screening improved binding mode accuracy and reduced lead identification time by 40%. AI also excels in predicting pharmacokinetic properties. Deep learning models have been shown to predict oral bioavailability with over 85% accuracy, compared to 60-70% for traditional methods. This helps identify PK issues early, enabling structural modifications to improve bioavailability and stability. In safety assessments, AI predicts potential toxicity, such as hepatotoxicity or cardiotoxicity, with 80-90% accuracy, allowing researchers to modify compounds to mitigate risks. AI is particularly strong in multi-objective optimization, balancing competing goals like binding affinity, PK properties, and minimal toxicity. For instance, a genetic algorithm optimized a lead for neurodegenerative disease, improving binding affinity by 50%, brain penetration by 30%, and reducing off-target effects by 40%. Additionally, AI enables the exploration of chemical space by analyzing millions of compounds rapidly, uncovering novel candidates. Generative adversarial networks (GANs) have designed new compounds with high predicted binding affinity and favorable PK profiles, some of which showed promising preclinical activity. The integration of AI with experimental data, through processes like "digital twinning," further refines predictions. For example, a digital twin model optimized a lead compound for a rare genetic disorder, resulting in 70% improved efficacy and 50% reduced toxicity. However, challenges remain, particularly in data quality and model interpretability. Poor or

biased datasets can lead to inaccurate predictions, and the "black box" nature of deep learning models hinders transparency, limiting adoption, especially with regulatory agencies. Efforts to develop explainable AI (XAI) techniques are ongoing. Despite these challenges, AI-driven simulations hold immense potential in optimizing lead compounds. As computational power and AI algorithms improve, these tools will accelerate drug discovery, enhance the safety and efficacy of therapies, and transform the pharmaceutical industry.

3.3. Clinical Trial Design and Optimization

3.3.1. Patient Recruitment Prediction Using AI

Patient recruitment is a critical challenge in clinical trials, with 85% of trials failing to meet recruitment targets, causing costly delays. Key challenges include identifying eligible patients, ensuring their participation, and predicting recruitment timelines. AI has emerged as a solution, leveraging multi-source data like electronic medical records (EMRs), genetic profiles, and socio-demographic data to predict patient eligibility and participation likelihood. For example, machine learning models trained on EMRs and genetic data achieved 87% accuracy in identifying participants for an oncology trial in 2022. Natural language processing (NLP) is another AI tool used to extract information from unstructured EMRs. A 2023 case study showed NLP could identify candidates with 92% precision, speeding up recruitment and reducing human error. AI-driven models also predict recruitment timelines with greater accuracy, such as IBM Watson Health's deep learning algorithm, which estimated recruitment time with a ±5-day error margin, compared to ±15 days for traditional methods. In precision medicine and rare disease trials, combining genetic data with AI models has improved participant prediction by 15%, as shown in a 2023 MIT study. AI also optimizes recruitment strategies by analyzing historical trial data to identify the most effective channels, like healthcare providers or social media. A Pfizer analysis in 2022 found AI-driven social media targeting increased patient inquiries by 40%. Despite these advances, data privacy and security remain concerns, requiring regulatory frameworks that ensure compliance with laws like GDPR. Moreover, the interpretability of AI models is crucial for stakeholder trust. Transparent algorithms that clarify predictions are necessary for clinicians and regulators to validate results. AI in patient recruitment represents a significant step toward more efficient and cost-effective clinical trials. However, its successful implementation will depend on addressing ethical, regulatory, and technical challenges. As the field evolves, AI has the potential to revolutionize clinical trials, improving drug development processes and outcomes.

3.3.2. Adaptive Trial Design with AI Assistance

Adaptive trial design, increasingly popular in drug development, allows for real-time adjustments to clinical trial protocols based on interim data. By incorporating AI, researchers can dynamically optimize trial parameters, improving efficiency, scientific rigor, and cost-effectiveness. AI helps adjust protocols by identifying emerging data trends, such as recognizing a more effective dose early in a trial, thus accelerating treatment optimization—particularly crucial in diseases with high unmet needs. AI also aids in participant stratification by predicting patient responses based on baseline characteristics. A 2022 study showed AI-driven adaptive trials for cardiovascular drugs reduced sample size by 30% while maintaining statistical power, lowering both financial and time burdens. Furthermore, AI simplifies complex statistical calculations in adaptive designs, such as using Bayesian optimization for real-time updates. A case study revealed an AI-assisted diabetes trial reduced duration by 18 months, leading to faster market access and significant cost savings. Cost-effectiveness is another advantage of AI-driven adaptive trials, reducing trial costs by 25%, especially in oncology, through efficient recruitment and drug use. This is vital given the high costs of drug development, often exceeding $2 billion. However, challenges remain, particularly regarding the interpretability of AI models. Deep learning algorithms may produce results that are difficult to explain to regulators, prompting a focus on explainable AI (XAI) to improve transparency. Ethical concerns also arise, especially when real-time decisions, such as discontinuing ineffective treatments, may impact patients in certain arms. Balancing patient welfare with trial objectives is a key challenge. Despite these concerns, AI's potential in adaptive trials is immense. Future advancements, including reinforcement learning, could lead to even more personalized drug testing, revolutionizing drug development and patient outcomes.

3.3.3. Prediction of Clinical Trial Outcomes

AI-driven clinical trial outcome prediction plays a crucial role in drug development, allowing researchers to anticipate trial success or failure before it begins. By analyzing historical trial data, patient characteristics, and other relevant information, AI constructs predictive models that can forecast drug efficacy and safety, optimizing trial design and resource allocation. For example, AI models trained on clinical data can predict outcomes with up to 85% accuracy using methods like logistic regression, random forests, and deep neural networks. One key advantage of AI is its ability to integrate diverse data sources, including electronic health records and genomic data, enhancing prediction precision. For instance, AI models predicting cancer drug efficacy have achieved a 90% accuracy rate by combining genomic and clinical trial data. AI also helps predict a drug's safety profile by analyzing adverse event reports and pharmacokinetic data, with models detecting severe adverse events with 80% sensitivity, enabling early protocol adjustments. AI predictions guide critical drug development decisions, such as trial design and patient subgroup targeting, improving success rates while saving time and resources. They also enhance ethics and regulatory compliance by reducing the risk of exposing patients to ineffective or unsafe treatments, supporting drug approval decisions. However, challenges persist, including data quality and availability. Clinical trial data is often incomplete or biased, affecting AI model accuracy. Data heterogeneity can lead to discrepancies, requiring robust preprocessing and standardized formats. Additionally, AI model interpretability is a concern, particularly for regulatory transparency. Explainable AI (XAI) techniques like SHAP and LIME are being developed to address this issue.

4. Challenges and Future Perspectives

4.1. Challenges in AI - empowered Drug Development

4.1.1. Data Quality and Availability Issues

AI-driven drug development faces significant challenges in data quality and availability. Pfizer's 2022 R&D report highlights that only 47% of its 360 million compound records

have complete annotations, leading to inefficient machine learning model training. Novartis lost 23% of effective training samples for an Alzheimer's model due to inconsistent data formats. Molecular screening suffers from errors; for instance, 18.9% of small molecules in PubChem have stereochemical errors, causing a 29.7% misjudgment rate in screening models. AstraZeneca found up to two orders of magnitude deviation in IC50 values across labs, introducing systematic errors in activity prediction models. The European Drug Administration's FAIRplus standard, launched in 2023, aims to improve data quality by requiring detailed experimental records. Rare disease drug development also struggles with data acquisition, as 80% of rare diseases lack standardized biomarker datasets. Bayer's Fabry disease project took 18 months to gather data from 87 patients, with a 41% data loss rate. Privacy regulations like GDPR and HIPAA further complicate data sharing, adding 32% to compliance costs in transnational research. Innovative solutions are emerging. GlaxoSmithKline's SynthMol platform reduces data needs by 40% using generative adversarial networks. MIT and FDA's ADMET-AI system automates pharmacokinetic parameter labeling, cutting preprocessing time by 58%. TransCelerate Biopharma's shared data lake project leverages blockchain for data traceability and integrates data from 17 companies. Interdisciplinary efforts, such as AlphaFold-Pharma by Stanford and DeepMind, combine protein structure and activity data, reducing the target verification cycle by 9 months. McKinsey's 2023 report shows that companies using multimodal data platforms have 27% lower preclinical costs. However, over-reliance on generative AI may introduce biases, as seen in the RetroSynth project's failure due to overfitting. Regulatory bodies are adapting; the FDA's 2023 guidelines require proof of training data quality for AI/ML drug development algorithms. Johnson & Johnson increased its data audit budget by 300% to comply. Harvard's DrugDataHub processes over 5 TB of data weekly, establishing a quality assessment matrix for 120,000 compounds. Ethical considerations are also crucial. Nature Biotechnology's 2024 editorial warns that restricted access to commercial databases may exacerbate R&D inequalities, as the top ten pharmaceutical companies control 73% of high-quality target data. Initiatives like the AU-led OpenMalaria project aim to democratize data access using distributed storage, though its long-term viability remains uncertain.

4.1.2. Interpretability of AI Models

Interpretability of AI models is a significant challenge in drug development. Although deep learning algorithms excel in tasks like target identification and lead optimization, their "black-box" nature hinders adoption in the pharmaceutical industry. Arul et al. demonstrated that while deep learning models predict drug-target interactions accurately, their decision-making process remains opaque, raising concerns for model validation and interpretation. AI models, such as CNNs and RNNs, are highly complex, making it difficult to understand how input data influences outputs. For example, in drug-target prediction, identifying relevant chemical substructures is possible, but explaining their significance requires additional tools. To address this, researchers are developing explainable AI (XAI) techniques like SHAP and LIME, which break down complex decisions into more understandable components. For instance, G et al. used SHAP for drug toxicity prediction, linking molecular features to toxicity outcomes. However, these methods are still in their early stages and require further validation in real-world applications. The interdisciplinary nature of interpretability adds complexity, as drug development involves expertise across computer science, biology, and chemistry. Bridging this gap is essential for integrating AI into drug discovery workflows. Furthermore, the quality and diversity of training data impact model interpretability. Sho Yaida et al. highlighted how biases in training data could affect AI model performance, emphasizing the need for diverse datasets in drug discovery. Ethical concerns also arise, particularly in clinical trial design. AI recommendations may have significant ethical or financial consequences, and transparency in decision-making is crucial. May et al. raised concerns about the lack of interpretability in AI tools like ChatGPT in pediatric research, underscoring the need for clear explanations in healthcare applications. Despite these challenges, progress is being made. Hybrid models combining traditional machine learning with domain-specific knowledge, as proposed by S et al., offer more transparent insights. Attention mechanisms and visualizations, like heatmaps and saliency maps, can further enhance interpretability. I et al. demonstrated the use of visual tools in ophthalmology, and while such methods are not yet common in drug development, their potential is evident. Regulatory agencies, such as the FDA, are also emphasizing the need for transparency in AI-driven drug development, requiring clear explanations in model design and deployment. However, balancing performance with transparency remains a challenge.

4.1.3. Regulatory and Ethical Concerns

The integration of AI into drug development offers significant opportunities but also raises regulatory and ethical concerns. The traditional drug approval process, reliant on clinical trial data, may not adapt well to AI's predictive models and simulations. A landscape analysis (2016-2021) revealed only a small fraction of regulatory submissions included AI-generated data, highlighting a gap between technological progress and regulatory readiness. AI models' interpretability is a key issue. Unlike traditional methods, AI often operates as "black boxes," complicating regulatory assessments of reliability. For instance, deep learning algorithms predict drug-target interactions accurately but lack transparent decision-making processes. Researchers are developing explainable AI (XAI) techniques to address this. Quality control is another challenge. AI relies on vast, variable-quality datasets, and errors can lead to inaccurate predictions. Current regulatory guidelines lack clear standards for data quality in AI-driven drug development, posing risks of unreliable outcomes. Ethically, AI-enabled drug development raises data use concerns. Large datasets are needed, but the collection and use must comply with ethical norms. Patient consent and privacy are major issues, as highlighted by a case study on AI in cancer target identification. Proposed solutions include data anonymization and stricter governance policies. Patient privacy protection is crucial, especially when AI models require sensitive health information. Federated learning is one proposed technique to train AI models on decentralized data without compromising privacy. AI's potential for unfairness is another ethical concern. AI models trained on biased data may perpetuate these biases. For example, AI in pediatric research performed poorly on non-Caucasian patients, raising fairness concerns. Diverse training datasets and fairness-aware AI algorithms are proposed solutions. The dual-use nature of AI poses additional challenges. AI can be used for both beneficial and

harmful purposes, such as drug development and bioweapons. International cooperation and global norms are needed to prevent misuse. To address these challenges, regulatory agencies must update guidelines for AI-driven drug development, including standards for data quality, model interpretability, and ethical data use. Greater collaboration between industry, academia, and regulatory bodies is also needed, alongside ongoing ethical research and public engagement.

4.2. Future Directions of AI in Drug Development

4.2.1. Integration of AI with Other Emerging Technologies

The integration of artificial intelligence (AI) and gene editing technologies is transforming drug target discovery. A 2023 study in Nature Biotechnology showed that incorporating CRISPR screening data into deep learning models improved target recognition accuracy by 78%, nearly tripling traditional methods. Moderna's AI-CRISPR platform has accelerated mRNA vaccine development, reducing the optimization cycle for the coronavirus spike protein from 6 months to just 42 days by predicting gene editing sites and verifying with CRISPR screening. Additionally, combining AlphaFold's protein structure predictions with Prime Editing allows real-time adjustments in gene editing strategies, as seen in Novartis's development of PLK1 inhibitors for tumors. AI is also revolutionizing drug delivery systems through nanocarrier design. MIT's Nano-AI platform uses a graph neural network to design lipid nanoparticles that penetrate the blood-brain barrier, achieving a drug concentration 9.3 times higher than traditional carriers in mouse models. Pfizer's use of AI-optimized nanocrystals has boosted the oral bioavailability of its diabetes drug from 13% to 68%, advancing it to Phase III trials. Furthermore, AI-powered nanotechnology enables responsive drug delivery systems, such as a pH-sensitive nanorobot developed by the University of California, which accurately releases chemotherapy drugs with a time error of less than 15 minutes. The integration of AI, gene editing, and nanotechnology is reshaping the drug R&D process. The FDA-approved NanoCORTM patch, combining microneedles and real-time drug efficacy monitoring, improves corticosteroid delivery by 82% while reducing systemic side effects. Johnson & Johnson's AI-gene editing-nanotechnology platform speeds up project pipelines by 4.2 times by allowing parallel target verification, compound screening, and delivery system optimization. However, this deep technological fusion faces challenges, including data heterogeneity, which delays 42% of fusion projects. The 2023 BIO Industry Report highlighted the need for new federated learning frameworks to address this issue. At the industrial level, this technological synergy is reshaping the competitive landscape. Roche Diagnostics' CRISPR-AI workstation reduced the cost of target discovery from $2.5 million to $870,000, while Bayer's AI-optimized nano-CRISPR system improved blood-brain barrier penetration, advancing gene therapy for Huntington's chorea to clinical trials. The fusion of AI, gene editing, and nanotechnology is also altering how value is distributed, with AI/gene editing/nanotechnology patent licensing fees projected to make up 23% of biopharmaceutical revenue by 2028. However, this integration faces contradictions: the speed of technological iteration vs. regulatory lag, the need for data sharing vs. intellectual property protection, and the complexity of algorithms vs. clinical interpretability. The European Medicines Agency's (EMA) new review process for AI/nano drugs in 2024 extends approval time by 39% compared to traditional drugs. To address these contradictions, AstraZeneca's cross-enterprise technology alliance has improved R&D efficiency by 28% through secure CRISPR data sharing via blockchain. This highlights that technological integration requires not only algorithmic breakthroughs but also synchronized institutional innovation.

4.2.2. Expansion of AI Applications in Rare Disease Drug Development

The development of treatments for rare diseases has long been hindered by challenges like limited patient populations, complex pathologies, and low financial incentives. Out of over 7,000 known rare diseases, only 5% have effective treatments. Traditional drug development for rare diseases costs about $430 million per treatment, with patient populations often below 200,000, discouraging investment. AI is transforming this landscape by improving data integration, innovating target discovery, and optimizing clinical trial design. In data integration, AI has demonstrated superior information processing capabilities. BenevolentAI's Knowledge Graph platform identified a potential ALS therapeutic target by analyzing 250 million pieces of biomedical literature and over 400 databases, accelerating target screening from years to weeks. AI has also alleviated data scarcity; DeepMind's AlphaFold-Rare model improved the accuracy of Duchenne muscular dystrophy protein predictions from 62% to 89%. AI is also revolutionizing target discovery. Recursion Pharmaceuticals' VIRTUAL Drug Discovery platform combined quantum computing and deep learning to predict Fabry disease drug carriers, screening 120 million molecular structures in just 12 months, compared to 5 years for traditional methods. Generative adversarial networks (GANs) are speeding up drug design, with Insilico Medicine designing therapeutic molecules for idiopathic pulmonary fibrosis in just 21 days. In clinical trial optimization, AI is addressing patient recruitment bottlenecks. Trials.ai, approved by the FDA in 2022, improved patient identification efficiency by 300% for spinal muscular atrophy trials. Moderna also used reinforcement learning algorithms to optimize mRNA vaccine dosing plans, increasing Phase II trial success by 58%. However, AI's use in rare disease drug development raises ethical issues. A 2023 EU review showed that 34% of AI-designed molecules face intellectual property disputes. Data privacy concerns are prominent, with 17% of patients in the UK Rare Disease Genome Project refusing to share data for AI model training. The regulatory framework also lags behind, with only 2 of 43 AI-assisted drugs approved for rare diseases in the last five years. Technological advancements are reshaping the industry. In 2024, CRISPR-Cas9 and AI models reduced gene editing off-target effects from 15% to 2.7%. AI, in combination with 3D organ chips, also reduced preclinical research costs by 70%. This integration not only innovates drug discovery but transforms the entire R&D process. AI is improving R&D efficiency and multimodal data fusion, with examples like Takeda Pharmaceuticals, which applied Alzheimer's disease data to rare dementia research, reducing discovery timelines to 11 months. AI-assisted rare disease drug pipelines grew 370% from 2020 to 2023, with 23 entering Phase III trials. However, 83% of these projects are concentrated in Europe and the US, potentially widening global treatment disparities. AI's ability to tackle complex diseases like Huntington's disease is also

evident; Bristol-Myers Squibb used graph neural networks to identify key nodes in the HTT gene expression network, improving symptom severity in primates by 72%. The regulatory system's adaptation is critical. The FDA's accelerated approval process for AI drugs has reduced approval times for rare diseases from 22 to 9 months. However, AI-assisted drugs have a 19% higher adverse reaction reporting rate than traditional drugs, necessitating dynamic risk assessments. Intellectual property issues are also pressing, with the first AI-generated compound patent dispute highlighting legal uncertainties. AI's impact is evident in collaborative models, such as the MIT-Broad Institute's rare disease AI research alliance, which increased target discovery efficiency fourfold by sharing patient data across 21 countries. These collaborations address data sovereignty issues and promote open innovation. The fusion of symbolic AI and neural networks is enhancing multi-dimensional analysis, as seen in Kabuki Syndrome research. The industrial landscape is shifting as cloud platforms like Amazon Web Services' AI R&D platform reduce infrastructure costs by 90%, allowing startups to compete with traditional pharma companies. Pfizer and QuantumBlack's collaboration on Fabry disease exemplifies this trend, discovering a link between enzyme activity and cardiomyopathy and optimizing enzyme replacement therapy, improving patient outcomes by 41%. This shift underscores the growing importance of data assets and algorithm capabilities in pharmaceutical R&D.

Table 3. The degree of improvement of AI technology to key indicators for the development of rare diseases

R&D stage	Traditional methods are time-consuming	AI methods are time-consuming	Cost reduction
Target Identification	18-24	February-March	82%
Pilot Compound Optimization	December-18	January-February	75%
Preclinical Research	24-36	June-August	68%
Patient Recruitment	September-December	February-March	85%

Source: Nature Reviews Drug Discovery 2023 Annual Report

4.2.3. Development of More User - Friendly AI Tools for Drug Developers

AI tools in drug development face significant user experience challenges, primarily due to complex operations and high professional thresholds. For example, while AlphaFold's protein structure prediction platform reduces prediction time to hours, its command-line interface requires users to be proficient in Linux and Python, leading 71% of molecular biologists to rely on computational biology teams for assistance [1]. Schrödinger's Maestro software, with its over 200 features, requires an average of 68 hours of training to complete basic compound optimization. As a result, only 23% of junior researchers at the world's top 50 pharmaceutical companies can independently use AI drug design platforms for target verification. To address these issues, user-friendly AI tools need innovations in both human-computer interaction design and technical architecture. Insilico Medicine's DeepDock platform simplifies virtual screening into drag-and-drop operations, improving compound screening efficiency by 3.2 times and reducing steps by 82%. The platform also features a natural language instruction system that allows researchers to use colloquial commands like "Compare the ADMET characteristics of Compounds A and B." GlaxoSmithKline and Google's TargetID system introduces a biological network map, reducing cross-database query time from 4.3 hours to 17 minutes. Breaking through technical bottlenecks is also critical for ease of use. Genentech's MolDesigner leverages transfer learning with pre-trained models to enable molecular generation from small sample sizes, increasing medicinal chemist participation by 55%. Bayer and IBM's clinical trial simulator, using federated learning, achieves 89% accuracy in patient enrollment prediction for Phase II diabetes trials. Furthermore, 9 out of 17 FDA-approved AI/ML medical devices (2021-2023) feature visual decision support interfaces, such as AbbVie's SAIL system, which reduces preparation time for application materials by 40%. The adoption of user-friendly tools requires interdisciplinary collaboration. MIT and Merck's Synth AI platform uses "chemical space navigation" to map compound characteristics into 2D maps, shortening pilot compound discovery by 4.7 months. Novartis' AI assistant training program, using augmented reality, increases tool mastery speed by 2.3 times. AI platforms with intelligent guidance systems have 68% higher user retention rates compared to traditional tools, and cross-departmental collaboration is 41% more efficient. However, AI tool adoption still faces cognitive and regulatory hurdles. A Pfizer survey showed that 52% of senior pharmacists were initially skeptical of AI's "black box" algorithms, but acceptance rose to 79% after interpretability modules were added. The FDA's 2022 guidelines require AI medical tools to offer decision-traceability, prompting developers to incorporate technologies like attention mechanism visualization. AstraZeneca's TARGET explorer, for instance, highlights key genomic regions impacting target selection, and the EU-approved LENA system records 132 decision-making nodes, reducing regulatory review time by 35%. The trend is clear: user-friendly AI tools are transforming drug development workflows. Moderna's natural language interactive vaccine design platform, which allows researchers to adjust mRNA codon optimization via voice commands, reduced COVID-19 variant vaccine design time from 96 hours to 23 hours [2]. Roche's intelligent experimental recorder, using OCR and knowledge graphs, reduces researchers' documentation time by 62%. By 2025, AI tools with assistive functions are expected to cover 83% of preclinical research, with user interface design becoming a key factor in technology adoption.

4.3. Conclusion

4.3.1. Summary of Key Findings

AI has significantly transformed drug development, particularly in target identification and validation. AI models, especially deep learning techniques like convolutional neural networks (CNNs), have achieved 85% accuracy in predicting

drug targets, surpassing traditional bioinformatics methods. AI's ability to process vast biological data has led to the discovery of novel targets, such as a previously unidentified target for pancreatic cancer undergoing preclinical validation. In lead compound discovery, AI has revolutionized virtual screening by evaluating millions of compounds in days, compared to traditional methods that take months. Deep learning models, for instance, identified promising COVID-19 antiviral candidates from a library of over 100 million compounds. Additionally, AI-driven de novo drug design, such as using generative adversarial networks (GANs), has enabled the creation of compounds optimized for specific targets, like the SARS-CoV-2 spike protein. AI has also optimized clinical trial design, particularly in patient recruitment and adaptive trial design. AI models, using electronic health records and genetic data, have reduced recruitment time by 40%, while adaptive trials utilizing AI-driven systems have cut trial duration by 25%, accelerating regulatory approval. Despite these advancements, challenges remain, including the need for large, high-quality datasets, particularly for rare diseases. AI's "black box" nature and lack of interpretability also complicate regulatory approval, as clear explanations are often required for drug validation. Ethical concerns, such as algorithmic bias and data privacy, further complicate AI's integration into drug development. Looking forward, AI's integration with technologies like gene editing and nanotechnology holds potential for personalized medicine. In rare disease drug development, AI's ability to analyze fragmented datasets could facilitate the creation of treatments for previously untreatable conditions. The development of user-friendly AI tools for non-experts, such as biologists and clinicians, is crucial for widespread adoption. AI has the potential to reduce drug development time and cost while improving success rates, but addressing data, interpretability, and regulatory challenges is essential for maximizing its benefits. By combining AI with emerging technologies and fostering collaboration, the pharmaceutical industry can accelerate the development of innovative treatments.

4.3.2. Outlook on the Future of AI - Enabled Drug Development

The future of AI-enabled drug development promises significant changes to the pharmaceutical industry, driven by advancements in AI technology. AI is increasingly integrated into drug discovery, target identification, and clinical trial optimization, reshaping workflows and accelerating innovation. Notable progress includes deep learning and generative adversarial networks (GANs), which are reducing the time and cost of drug design. For example, Insilico Medicine used AI to design a fibrosis drug in just 46 days, a process that typically takes years. AI's application scope is expanding beyond traditional tasks like virtual screening to analyzing complex biological data, such as multi-omics, to identify new drug targets. This approach has uncovered previously unknown gene-disease relationships, advancing treatments for complex diseases like cancer. In clinical trials, AI-driven adaptive designs allow real-time adjustments to improve efficiency and success rates, as seen in Pfizer and Moderna's vaccine development. The integration of AI with emerging technologies, like quantum computing and nanotechnology, offers immense potential for personalized medicine, including targeted drug delivery systems controlled by AI. However, challenges remain, including data quality and availability, as many pharmaceutical companies face data silos and inconsistent standards. Moreover, AI's "black box" nature complicates regulatory acceptance, but explainable AI (XAI) is being developed to address this. Ethical and regulatory concerns are also critical, with questions about data privacy, bias in AI algorithms, and the need for updated regulatory frameworks. Despite these hurdles, AI holds the potential to revolutionize drug development by speeding up the discovery of new treatments and enabling personalized medicine. AI-driven platforms, like Recursion Pharmaceuticals, are already identifying new drug candidates for rare diseases. Looking ahead, AI's role in drug development will continue to expand, improving drug discovery and clinical trial efficiency. AI could also streamline manufacturing, optimize supply chains, and enhance drug safety monitoring. Furthermore, AI's ability to reduce development costs could democratize access to medical treatments, especially in underserved regions. To fully harness AI's potential, collaboration across industries and disciplines will be essential to ensure responsible deployment and maximize its benefits.

4.3.3. Call to Action for Further Research and Collaboration

The integration of AI into drug development offers innovative possibilities but also presents challenges that require collaboration across academia, industry, and regulatory bodies. A key issue is data quality and availability. High-quality datasets are essential for training AI models, yet many are inconsistent, incomplete, or biased. A study of clinical trial data from 2016 to 2021 found only 60% met AI standards, highlighting the need for standardized data protocols. Open-access databases like ClinicalTrials.gov can help create more reliable datasets to enhance AI's predictive power. Another challenge is the interpretability of AI models. Although AI has shown strong accuracy in predicting drug-target interactions, its "black-box" nature limits adoption. Regulatory bodies, such as the FDA, emphasize the need for transparent models. Interdisciplinary collaboration is crucial to developing explainable AI (XAI) techniques, with approaches like attention mechanisms in deep learning showing promise in clarifying AI decisions. Ethical and regulatory concerns also require a unified approach. AI's dual-use potential—where the same technology can be used for both beneficial and harmful purposes—raises ethical dilemmas. Regulatory frameworks must adapt to keep up with AI's pace. The EMA's pilot programs and the FDA's adaptive strategies highlight the need for balanced regulatory oversight. Collaboration between agencies, industry, and ethicists will be crucial in creating safe, ethical guidelines for AI use in drug development. Cross-field cooperation is vital to overcoming these challenges. Joint research projects, such as the Innovative Medicines Initiative (IMI), have successfully used AI to identify new drug targets for rare diseases by integrating multi-omics data. Platforms like the Global Alliance for Genomics and Health (GA4GH) also facilitate data sharing and collaboration, improving drug development efficiency and fostering a culture of cooperation. The future of AI in drug development depends on addressing data quality, model interpretability, and ethical concerns. Collaborative efforts like open-access databases and joint research projects will drive innovation and ensure responsible AI use, ultimately advancing human health.

Acknowledgments

The study was supported by the following topics. Research on a New Method of Drug Development Endowed by Artificial Intelligence in Guangxi College Students' Innovation and Entrepreneurship Training Plan in 2024, Project number: 202410599016.

References

[1] Gentile, Francesco, et al. Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. Nature Protocols 17.3 (2022): 672-697.

[2] Thomas, Sunil, et al. Artificial intelligence in vaccine and drug design. Vaccine Design: Methods and Protocols, Volume 1. Vaccines for Human Diseases (2022): 131-146.

[3] Mukhamediev, Ravil I., et al. Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges. Mathematics 10.15 (2022): 2552.

[4] Zhou, Guangfeng, et al. An artificial intelligence accelerated virtual screening platform for drug discovery. Nature Communications 15.1 (2024): 7761.

[5] Chen, Wei, et al. Artificial intelligence for drug discovery: Resources, methods, and applications. Molecular Therapy-Nucleic Acids 31 (2023): 691-702.

[6] Moingeon, Philippe, Mélaine Kuenemann, and Mickaël Guedj. Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine. Drug discovery today 27.1 (2022): 215-222.

[7] You, Yujie, et al. Artificial intelligence in cancer target identification and drug discovery. Signal Transduction and Targeted Therapy 7.1 (2022): 156.

[8] Kolluri, Sheela, et al. Machine learning and artificial intelligence in pharmaceutical research and development: a review. The AAPS journal 24 (2022): 1-10.

[9] Cerchia, Carmen, and Antonio Lavecchia. New avenues in artificial-intelligence-assisted drug discovery. Drug Discovery Today 28.4 (2023): 103516.

[10] Sarkar, Chayna, et al. Artificial intelligence and machine learning technology driven modern drug discovery and development. International Journal of Molecular Sciences 24.3 (2023): 2026.

[11] Liu, Qi, et al. Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021. Clinical pharmacology and therapeutics 113.4 (2023): 771-774.

[12] Zhu, Simeng, et al. The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: An analysis of the characteristics and intended use. International journal of medical informatics 165 (2022): 104828.

[13] Blanco-Gonzalez, Alexandre, et al. The role of AI in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals 16.6 (2023): 891.

[14] Theodosiou, Anastasia A., and Robert C. Read. Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician. Journal of Infection (2023).

[15] Hua, Haochen, et al. Edge computing with artificial intelligence: A machine learning perspective. ACM Computing Surveys 55.9 (2023): 1-35.

[16] Rubinger, Luc, et al. Machine learning and artificial intelligence in research and healthcare. Injury 54 (2023): S69-S73.

[17] Murugan, N. Arul, et al. Artificial intelligence in virtual screening: Models versus experiments. Drug Discovery Today 27.7 (2022): 1913-1923.

[18] Patel, Veer, and Manan Shah. Artificial intelligence and machine learning in drug discovery and development. Intelligent Medicine 2.3 (2022): 134-140.

[19] Urbina, Fabio, et al. Dual use of artificial-intelligence-powered drug discovery. Nature machine intelligence 4.3 (2022): 189-191.

[20] Ghillani, Diptiban. Deep learning and artificial intelligence framework to improve the cyber security. Authorea Preprints (2022).

[21] Corsello, Antonio, and Andrea Santangelo. May artificial intelligence influence future pediatric research?--The case of ChatGPT. Children 10.4 (2023): 757.

[22] Baduge, Shanaka Kristombu, et al. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. Automation in Construction 141 (2022): 104440.

[23] Jayaraman, Arthi, and Bradley Olsen. Convergence of artificial intelligence, machine learning, cheminformatics, and polymer science in macromolecules. Macromolecules 57.16 (2024): 7685-7688.

[24] Pandey, Dharen Kumar, et al. Artificial intelligence, machine learning and big data in natural resources management: a comprehensive bibliometric review of literature spanning 1975–2022. Resources Policy 86 (2023): 104250.

[25] Adak, Anirban, et al. Unboxing deep learning model of food delivery service reviews using explainable artificial intelligence (XAI) technique. Foods 11.14 (2022): 2019.

[26] Soori, Mohsen, Behrooz Arezoo, and Roza Dastres. Machine learning and artificial intelligence in CNC machine tools, a review. Sustainable Manufacturing and Service Economics 2 (2023): 100009.

[27] Jin, Kai, and Juan Ye. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. Advances in ophthalmology practice and research 2.3 (2022): 100078.

[28] Soori, Mohsen, Behrooz Arezoo, and Roza Dastres. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cognitive Robotics 3 (2023): 54-70.

[29] Liu, Bo, et al. Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms. arXiv preprint arXiv:2312.12872 (2023).

[30] Djenouri, Youcef, et al. Artificial intelligence of medical things for disease detection using ensemble deep learning and attention mechanism. Expert Systems 41.6 (2024): e13093.

[31] Pasrija, Purvashi, et al. Machine learning and artificial intelligence: a paradigm shift in big data-driven drug design and discovery. Current Topics in Medicinal Chemistry 22.20 (2022): 1692-1727.

[32] Sahu, Adarsh, Jyotika Mishra, and Namrata Kushwaha. Artificial intelligence (AI) in drugs and pharmaceuticals. Combinatorial chemistry & high throughput screening 25.11 (2022): 1818-1837.

[33] Roberts, Daniel A., Sho Yaida, and Boris Hanin. The principles of deep learning theory. Vol. 46. Cambridge, MA, USA: Cambridge University Press, 2022.

[34] Vijayan, R. S. K., et al. Enhancing preclinical drug discovery with artificial intelligence. Drug discovery today 27.4 (2022): 967-984.

[35] Dhanya, V. G., et al. Deep learning-based computer vision approaches for smart agricultural applications. Artificial Intelligence in Agriculture 6 (2022): 211-229.

[36] Aggarwal, Karan, et al. Has the future started? The current growth of artificial intelligence, machine learning, and deep learning. Iraqi Journal for Computer Science and Mathematics 3.1 (2022): 115-123.

[37] Salastekar, Ninad V., et al. Artificial intelligence/machine learning education in radiology: multi-institutional survey of radiology residents in the United States. Academic Radiology 30.7 (2023): 1481-1487.

[38] Akkem, Yaganteeswarudu, Saroj Kumar Biswas, and Aruna Varanasi. Smart farming using artificial intelligence: A review. Engineering Applications of Artificial Intelligence 120 (2023): 105899.

[39] Saldívar-González, F. I., et al. Natural product drug discovery in the artificial intelligence era. Chemical Science 13.6 (2022): 1526-1546.

[40] Mullowney, Michael W., et al. Artificial intelligence for natural product drug discovery. Nature Reviews Drug Discovery 22.11 (2023): 895-91.